Chapter 7

Estimating functionals, I

From now on we switch from the problem of estimating a nonparametric regression function to the problem of estimating functional of such a function.

7.1 The problem

We continue to work within the bounds of the \( L_2 \)-theory and Gaussian white noise model of observations. Geometrical setting of the generic problem we are interested in is as follows:

We are given

- a real separable Hilbert space \( H \) with inner product \( (\cdot, \cdot) \) and an orthonormal basis \( \{\phi_i\}_{i=1}^{\infty} \),
- a set \( \Sigma \subset H \),
- a real-valued functional \( F \) defined in a neighbourhood of \( \Sigma \).

A “signal” \( f \in \Sigma \) is observed in Gaussian white noise of intensity \( \varepsilon \), i.e., we are given a sequence of observations

\[
y_f^{l,\varepsilon} = \{y_f^{l,\varepsilon} \equiv (f, \phi_l) + \varepsilon \xi_l\},
\]

\( \{\xi_l\}_{l=1}^{\infty} \) being a collection of independent \( \mathcal{N}(0,1) \) random variables (“the noise”), and our goal is to estimate via these observations the value \( F(f) \) of \( F \) at \( f \).

As always, we will be interested in asymptotic, \( \varepsilon \to 0 \), results.

Recall that the model (7.1) is the geometric form of the standard model where signals \( f \) are functions from \( L_2[0,1] \), and observation is the “functional observation”

\[
y_f(x) = \int_0^x f(s)ds + \varepsilon W(x),
\]

\( W(x) \) being the standard Wiener process; in this “functional language”, interesting examples of functionals \( F \) are the Gateau functionals

\[
F(f) = \int_0^1 G(x, f(x))dx
\]
or

\[ F(f) = \int_0^1 \int_0^1 G(x_1, ..., x_k, f(x_1), ..., f(x_k)) dx_1 ... dx_k. \quad (7.4) \]

In this chapter we focus on the case of a smooth functional \( F \). As we shall see, if the parameters of smoothness of \( F \) "fit" the geometry of \( \Sigma \), then \( F(f), f \in \Sigma \), can be estimated with "parametric convergence rate" \( O(\varepsilon) \), and, moreover, we can build **asymptotically efficient**, uniformly on \( \Sigma \), estimates.

### 7.1.1 Lower bounds and asymptotical efficiency

In order to understand what "asymptotical efficiency" should mean, the first step is to find out what are limits of performance of an estimate. The answer can be easily guessed: if \( F(f) = (f, \psi) \) is a continuous linear functional, so that

\[ \psi = \sum_{i=1}^{\infty} \psi_i \phi_i, \{\psi_i = (\psi, \phi_i)\}_{i=1}^{\infty} \in \ell^2, \]

then seemingly the best way to estimate \( F(f) \) is to use the "plug-in" estimate

\[ \hat{F}(y) = \sum_{i=1}^{\infty} \psi_i y_i \phi_i = (f, \psi) + \varepsilon \sum_{i=1}^{\infty} \psi_i \xi_i \]

(the series in the right hand side converges in the mean square sense, so that the estimate makes sense); the estimate is unbiased, and its variance clearly is \( \varepsilon^2 \|\psi\|^2 \).

Now, if \( F \) is Fréchet differentiable in a neighbourhood of a signal \( f \in \Sigma \), then we have all reasons to expect that locally it is basically the same – to estimate \( F \) or the linearized functional \( \tilde{F}(g) = \tilde{F}(f) + (F'(f), g - f) \), so that the variance of an optimal estimate in this neighbourhood should be close to \( \varepsilon^2 \|F'(f)\|^2 \). Our intuition turns out to be true:

**Theorem 7.1.1** [13] Let \( \bar{f} \in \Sigma \) and \( F \) be a functional defined on \( \Sigma \). Assume that

(i) \( \Sigma \) is convex, and \( F \) is Gateau differentiable "along \( \Sigma \)" in a neighbourhood \( U \) of \( \bar{f} \) in \( \Sigma \): for every \( f \in U \), there exists a vector \( F'(f) \in H \) such that

\[ \lim_{t \to 0} \frac{F(f + t(g - f)) - F(f)}{t} = (F'(f), g - f) \quad \forall g \in \Sigma, \]

and assume that every one of the functions \( \psi_g(t) = (F'(\bar{f} + t(g - \bar{f})), g - \bar{f}), \in \Sigma \), is continuous in a neighbourhood of the origin of the ray \( \{t \geq 0\} \)

(ii) The "tangent cone" of \( \Sigma \) at \( \bar{f} \) – the set

\[ T = \{h \in H \mid \exists t > 0 : \bar{f} + th \in \Sigma\} \]

is dense in a half-space \( H_+ = \{h \in H \mid (\psi, h) \geq 0\} \) associated with certain \( \psi \neq 0 \).

Then the local, at \( \bar{f} \), squared minimax risk of estimating \( F(f), f \in \Sigma \), via observations (7.1) is at least \( \varepsilon^2 (1 + o(1)) \|F'(\bar{f})\|^2 \):

\[ \lim_{\delta \to 0} \liminf_{\varepsilon \to 0} \inf_{\hat{F} \in F} \sup_{f \in \Sigma, ||f - \bar{f}|| \leq \delta} \mathcal{E} \left\{ \varepsilon^{-2} \left[ \hat{F}(y^{f,\varepsilon}) - F(f) \right]^2 \right\} \geq \|F'(\bar{f})\|^2, \quad (7.5) \]